Bayes’ Theorem

Bayes’ Theorem is written as:

P(A / B) = [P(B / A) * P(A)] / P(B)

where:

- **P(A / B)** is the probability of event A occurring given that event B has occurred.
- **P(B / A)** is the probability of event B occurring given that event A has occurred.
- **P(A)** is the prior probability of event A.

This can also be interpreted as: P(A / B) = [P(B / A) * P(A)] / [P(B / A) * P(A) + P(B / No A) * P(No A)]

It calculates the probability of event A given that event B has occurred, using the known conditional probability of B given A.

Proof of Bayes’ Theorem

We start with the definition of conditional probability:

1. **P(A / B) = P(A and B) / P(B)**
2. **P(B / A) = P(A and B) / P(A)**

From equation (2), we can write:

P(A and B) = P(B / A) * P(A)

Now substitute this into equation (1):

P(A / B) = [P(B / A) * P(A)] / P(B)

This is Bayes' Theorem.

Example: Medical Test Problem

Suppose a disease affects 1% of the population. A test for the disease has the following characteristics:

- If a person **has** the disease, the test is **positive** 99% of the time → P(Positive / Disease) = 0.99
- If a person **does not have** the disease, the test is **positive** 5% of the time → P(Positive / No Disease) = 0.05
- The chance of having the disease → P(Disease) = 0.01
- The chance of not having the disease → P(No Disease) = 0.99

A person takes the test and gets a **positive** result. What is the probability they actually have the disease?

Let:

- A = Disease
- B = Positive test

We apply Bayes' Theorem:

P(Disease / Positive) = [P(Positive / Disease) * P(Disease)] / P(Positive)

First, calculate P(Positive):

P(Positive) = P(Positive / Disease) * P(Disease) + P(Positive / No Disease) * P(No Disease)

P(Positive) = 0.99 * 0.01 + 0.05 * 0.99

P(Positive) = 0.0099 + 0.0495 = 0.0594

Now calculate P(Disease / Positive):

P(Disease / Positive) = 0.0099 / 0.0594 ≈ 0.1667

Conclusion: Even though the test is 99% accurate, the probability that a person actually has the disease given a positive result is only **16.67%**. This is because the disease is very rare, and false positives affect the result significantly.